SUBTASK 1

Given:

Empirical loss funcn for logistic regression is defined as:

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log (h_\theta(x^{(i)})) + (1-y^{(i)}) \log (1-h_\theta(x^{(i)}))\right]$$

$$y^{(i)} \in \{0,1\}$$

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}} = \sigma(z)$$

T.P.T:

Hessian matrix H of empirical loss function wrt θ is positive semidefinite in nature.

Proof:

Imp property to be used: $1-\sigma(z) = 1 - \frac{1}{1+e^{-z}} = \frac{e^{-z}}{1+e^{-z}}$

$$= \frac{1}{1+e^{z}} = \sigma(-z)$$

$$\therefore 1-\sigma(z) = \sigma(-z) \longrightarrow \text{Property 1.}$$

Imp. property 2: $\sigma'(z) = \sigma(z)(1-\sigma(z)) \rightarrow$ Property 2

Proof: $\sigma'(z) = \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}}\right) = \frac{e^{-z}}{(1+e^{-z})^2} = \left(\frac{e^{-z}}{1+e^{-z}}\right)\left(\frac{1}{1+e^{-z}}\right)$

$$= \sigma(z)(1-\sigma(z))$$

Hessian of $J(\theta) = \nabla^2 J(\theta)$

Note: Here, we will ignore $(-\frac{1}{m})$ factor for now for ease of calcns.

$$J(\theta) = -y^{(i)} \log h_\theta(x^{(i)}) \mp (1-y^{(i)}) \log (1-h_\theta(x^{(i)}))$$

$$= -y^{(i)} \log g(\theta^T x^{(i)}) \mp (1-y^{(i)}) \log (1-g(\theta^T x^{(i)}))$$

$$J'(\theta) = \nabla J(\theta) = \frac{\partial J}{\partial \theta^T} = -y^{(i)} \frac{\partial(\log g(\theta^T x^{(i)}))}{\partial \theta^T} - (1-y^{(i)}) \frac{\partial}{\partial \theta}(\log(1-g(\theta^T x^{(i)})$$

computing,

$$\frac{\partial \log g(\theta^T x^{(i)})}{\partial \theta^T} = \frac{1}{g(\theta^T x^{(i)})} \cdot \frac{\partial (g(\theta^T x^{(i)}))}{\partial \theta^T}$$

By chain rule:

$$= \frac{1}{g(\theta^T x^{(i)})} \cdot \frac{d(g(\theta^T x^{(i)}))}{d(\theta^T x^{(i)})} \cdot \frac{d(\theta^T x^{(i)})}{d\theta^T}$$

$$= \cancel{\frac{1}{g(\theta^T x^{(i)})}} \cdot \underbrace{\cancel{g(\theta^T(x^{(i)}))} \cdot (1-g(\theta^T x^{(i)}))}_{\text{By property 2}} \cdot x^{(i)}$$

$$= (1-g(\theta^T x^{(i)}))\, x^{(i)}$$

Whereas.

$$\frac{\partial \log (1-g(\theta^T x^{(i)}))}{\partial \theta^T} = \frac{1}{1-g(\theta^T x^{(i)})} \cdot \frac{\partial(1-g(\theta^T x^{(i)}))}{\partial \theta^T}$$

$$= \frac{1}{1-g(\theta^T x^{(i)})} \cdot \frac{-dg(\theta^T x^{(i)})}{d(\theta^T x^{(i)})} \cdot \frac{d(\theta^T x^{(i)})}{d\theta^T}$$ (chain rule)

$$= \cancel{\frac{1}{1-g(\theta^T x^{(i)})}} \cdot \underbrace{-g(\theta^T(x^{(i)})) \cdot \cancel{(1-g(\theta^T x^{(i)})}}_{\text{By property 2}} \cdot x^{(i)}$$

$$= -g(\theta^T(x^{(i)})) \cdot x^{(i)}$$

$$\therefore \vec{\nabla}^2 J(\theta) = \frac{\partial J(\theta)}{\partial\theta\, \partial\theta^T} = \frac{\partial}{\partial\theta}\left[\frac{\partial J(\theta)}{\partial\theta^T}\right]$$

$\frac{\partial J(\theta)}{\partial \theta^T}$ = using previously calculated values

$$= -x^{(i)} y^{(i)} (1-g(\theta^T x^{(i)})) + (1-y^{(i)})\, x^{(i)} (g(\theta^T x^{(i)}))$$

$$= x^{(i)} (g(\theta^T x^{(i)}) - y^{(i)})$$

$$\vec{\nabla}^2 J(\theta) = \frac{\partial}{\partial\theta}(\vec{\nabla} J(\theta)) = \frac{\partial}{\partial\theta}\left[x^{(i)}(g(\theta^T x^{(i)}) - y^{(i)})\right]$$

$$= x^{(i)} \cdot \frac{\partial(g(\theta^T x^{(i)}))}{\partial\theta}$$

By chain rule,

$$\vec{\nabla}^2 J(\theta) = x^{(i)} \cdot \frac{\partial (g(\theta^T x^{(i)}))}{\partial(\theta^T x^{(i)})} \cdot \frac{\partial (\theta^T x^{(i)})}{\partial \theta}.$$

$$= x^{(i)} \cdot \underbrace{g(\theta^T x^{(i)}) \cdot (1-g(\theta^T x^{(i)}))}_{\text{Property 2}} \cdot x^{T(i)}$$

$$\vec{\nabla}^2 J(\theta) = x^{(i)} x^{T(i)} g(\theta^T x^{(i)}) (1-g(\theta^T x^{(i)}))$$

$\therefore$ Hessian of $J(\theta) = -\frac{1}{m} x^{(i)} x^{T(i)} \cdot g(\theta^T x^{(i)}) (1-g(\theta^T x^{(i)}))$

PART - II : T.P.T. Hessian is a semidefinite matrix.

Cond^n: A matrix is said to be semidefinite positive iff.

(i) matrix is symmetric

(ii) Eigen values are non-negative.

i.e. $v^T M v \geq 0 \quad \forall v \in V$

Proof: (i) Matrix is symmetric

as partial second derivatives are equal.

i.e. eg. $f''_{xy} = f''_{yx}$ by Shwarz thm as $J(\theta)$ = continuous.

(ii) $\vec{v}^T M v \geq 0$.

$$\vec{v}^T [g(\theta^T x^{(i)}) (1-g(\theta^T x^{(i)})) x^{(i)} x^{T(i)}] \vec{v}$$

$$= g(\theta^T x^{(i)}) (1-g(\theta^T x^{(i)})) (\vec{v}^T x^{(i)}) (x^{T(i)} \vec{v})$$

$$= g(\theta^T (x^{(i)})) (1-g(\theta^T (x^{(i)}))) (\vec{v} x^{T(i)})^T (x^{T(i)} \vec{v})$$

$$= g(\theta^T x^{(i)}) (1-g(\theta^T x^{(i)})) \underbrace{(\vec{x}^{T(i)} \cdot \vec{v})^2}_{\text{squared value}} \geq 0.$$

This is always ≥ 0.

Both conditions are satisfied.

Hence proved.